

# 數學寶庫

## 的 SMILES 化學反應數據庫

### 鏈接

[價格信息表](#) [make\\_na](#)  
[server](#) [採購信息](#)  
[在 Wolfram](#) [功能網](#)  
[站](#) [法律公告](#) [生物](#)  
[物理軟件](#) [Smiles](#) [會員](#)  
[反應數據庫](#)  
[ChemAxon](#)  
[CFTR](#) [的基因組學](#)  
[基因治療網](#) [NCBI](#)  
[數據庫](#)  
[ChemSpider](#) [數據](#)  
[庫](#) [CFTR](#) [的維基](#) [基因](#)  
[與疾病](#) [Mathematica](#)  
[8 文件](#)

### 博客

### 數學寶庫

### 圖書預覽

### Gamma 函數

### 我們的書店

### Gamma 函數

有問題嗎？  
評論嗎？  
[發電子郵件給我們](#)

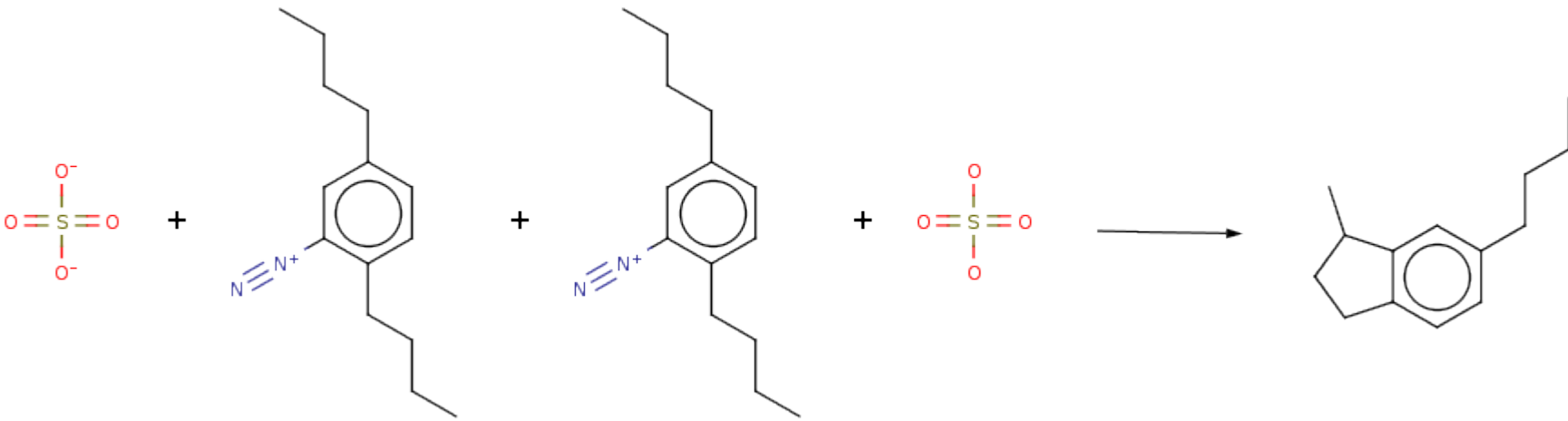
### 歡迎光臨

微笑的化學反應數據庫是一組包含反應物左右成對的結構信息和產品兩百萬不同的化學反應的文件。 簡化分子輸入行輸入系統的代表分子結構的（笑）是用來表示分子連接字符串的立體關係，的確化學反應。 這些的 SMILES 字符串表示形式啟發機器學習計算機程序，了解輸入/輸出關係之間存在反應的空間和產品空間，採用新穎的字符串變換算法（在本書的實施創造一種新的化學©2012，預計將公佈在 Amazon.com 在 2012 年秋季，使用 Mathematica 的編程語言）。

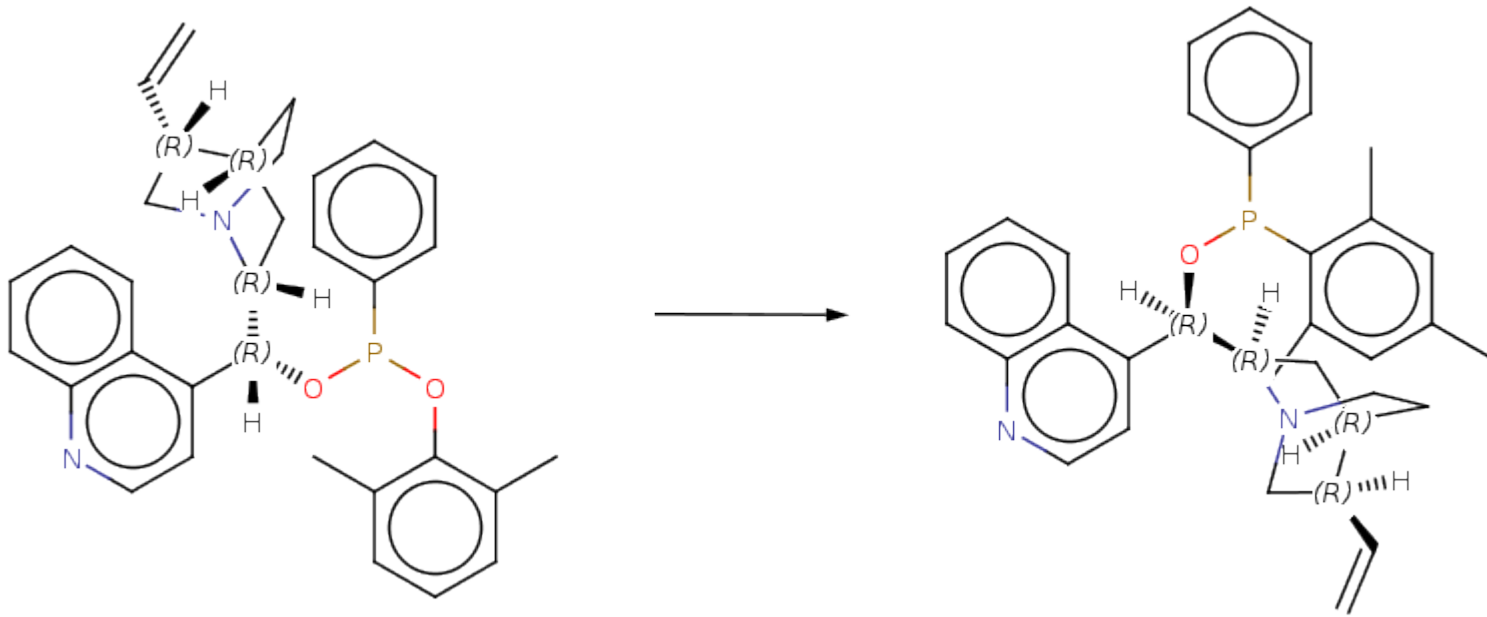
應用範圍：化學反應的結果預測，構效和 Retrosynthetic 分析。

作為示範使用字符串的微笑代表的連通性和立體幾何的化學結構和反應，和機器學習技術的實用，可以考慮以下兩個驗證這從一個數據集派生了一個數學模型正確預測 結果擁有 100,000 反應（其中被排除在外，這兩個反應）反應型材（結構和計量）有點類似（非常相似的情況下，測試的目的除外），小說每個測試用例：

[O-]S([O-])(=O)=O.CCCc1ccc(CCCC)c(c1)[N+]#N.CCCc1ccc(CCCC)c(c1)[N+]#N.OS(O)(=O)=O>>CCCCc1ccc2CCC(C)c2c1



[H][C@@](OP(Oc1c(C)cccc1C)c1ccccc1)(c1ccnc2ccccc12)[C@@]1([H])C[C@@]2([H])CCN1C[C@]2([H])C=C>>[H][C@@](OP(c1ccccc1)c1c(C)cc(C)cc1C)(c1ccnc2ccccc12)[C@@]1([H])C[C@@]2([H])CCN1C[C@]2([H])C=C



當然，機器學習技術同樣適用於 retrosynthetic 分析-心中有一個目標產品，一個是能夠預測成功的起點，事先合成的步驟材料的結構。許多暫定起始原料，或為一種人工合成的步驟線索，可以通過不同的預測模型計算得到，自己獲得每個新車型的基礎上的數據庫中的不同子集。這樣子集，可以選擇一些選擇標準，或隨機的，但在這種情況下，每次訓練的子集必須完全反應的反應物具有獨特的集，以避免多 僅的數據組成。

反應的預測是一個一（1:1）的關係，而 retrosynthetic 分析涉及一到多（1：M）的關係。在 retrosynthetic 分析的情況下，這種情況正在處理通過降低設置的地步，由此產生的模型，使一個良好的時間部分不正確的建議，訓練數據的大小。未納入一個重要的金額從數據庫中的知識（和可能的類型），該模型室發言創意的排序。然而，隨後通過一支訓練有素的反應預測模型的運行結果，我們借回確定性，從而確認是否建議的反應是可行的或不。

機器學習化學反應可以在三個非常重要的方式從比較正統的方法加以區分：第一，工作是完全不還原，不作為組成的亞原子粒子的行為的結果解釋化學反應，而是作為結果高等數學守恆定律。

要理解為什麼守恆定律，它代表數學的對稱性，考慮任何非共線數據點在直角平面。可能曲線可以通過這些數據點的數量是無限的。它是高度放肆，幾乎肯定錯誤的，天真地以為，連接數據點的平滑曲線代表正確任意彎曲的數據集的中間點。數據凝合，這在本質上，甚至包括如神經網絡技術本身，根本無法使用，數據籠統概括。事實仍然是，至少有一個條件，必須適用於曲線區分溶液的曲線。這需要一個模型的先驗知識。在任何形式的數據凝合，只有正確使用調整模型參數，而不是派生出一個模型。這是一個非常普遍的困擾，在計算智能領域的許多研究的監督。

我們在這項工作中，而不是尋找什麼是數學保守 比例因子。數學守恆定律 H 是同構的線性關係： $y=bx$ , 使得  $H(m(D_{i,2}))=\mu H(m(D_{i,1}))$  其中  $D_{i,j}$  是經驗分,  $\mu \neq 1$  是一個比例因子和  $m(\cdot)$  是化學指標。鑑於空間是離散和有限的，我們可以合法地結束，一個足夠簡單的函數 H 充分大的 i，有足夠大的我，和精心挑選的度量，數學守恆定律已經確定，並在值新穎點的  $[H(m(d_{r,1})),H(m(d_{r,2}))]_{\mu}$  之間的經驗分  $[H(m(D_{i,1})),H(m(D_{i,2}))]_{\mu}$  還在於沿直線連接線的經驗點。地圖可以被視為完成和  $d_{r,2}$  可以求解。整點的線性化是有  $\mathbb{N}_2$  可能不同的曲線，比實數集，  $\mathbb{N}_1$  的無限大。但綁定到一個特定點的線性射線是  $\mathbb{N}_1$ ，只取決於  $\theta$  的實際價值。

H 是通過一個進化的過程中搜查。 生隨機函數形式，通過兩輪交叉，變異，簡化和選擇。這兩個任務的性能和功能的簡單應用的選擇性壓力。簡單要求我們找到真正的保護功能。一個不合理的任務完成的功能效益為目標。

當我們應用我們的數學模型，建築技術的笑容反應數據庫或其中的任意子集的數學模擬，我們運用相同的邏輯化學空間的一個子集-所有的分子結構的離散空間。

第二個顯著的因素是高層次的數學守恆定律，我們用來預測反應後，直接根據：

- 實驗反應數據-反應數據庫存儲兩百萬的反應字符串。
- 獨特的化學圖形字符串表示形式 - SMILES。
- 獨特的，統一大小，為了依賴性和可逆的字符串的 產品矩陣（非交換）使用替代字符矩陣乘法的數學表示。
- 數據拼接-數據融合的定義，通過數學守恆定律的發現。
- 盡可能簡單的函數 H 的演變是關鍵。
- H 是一個標量函數，而 m 是一個矩陣函數。
- H 的功能形式是依賴於 m 的函數形式， $\mu$  值和  $D_{i,k}$ 。
- 化學指標-標量值的典型性的先進理論為基礎的矩陣函數。

由於字符串代表矩陣  $m(\cdot)$  是一個標量，我們基本上分配多維數據點，以點上的實線。這不會導致分配給多個多維數據點對單點實線。事實上，代表在飛機上所有點上的實線代表無限的大小之分的無限大小是相同的。因此，所有 n 維數據點獨特的作

業點上的實線是可能的，這是可證明的。 採取一個兩維平面上的點 (x,y)。 我們可以把我們將使用寫下 x 和 y 的數字，他們簡單地交錯。 此交錯技術成果，在每一個可能的點的實數和相同數量的平面地圖上沒有兩點。 這同樣的論點可以擴展到任意維數，只要我們有一個維度的數量有限。 維度的概念具有無限空間的大小或基數上沒有影響;尺寸是 cardinally 意義。 然而在這裡，我們正在處理，如果被認為是整個體積離散超體積，可數無窮，但在這種情況下-一個非常大的數量有限。 盡可能小的有機分子的總數，填充化學空間估計已超過 10<sup>60</sup>。 因此反應空間是不可思議的大，但有限的。

第三個顯著的因素是，機器學習技術，更明確，更高效，更能夠應用於化學反應的問題時比傳統的方法。 例如，一般只限於傳統的量子反應散射計算的準確度範圍內任何涉及少於 6 個原子的反應。 涉及超過 6 個原子反應散射問題成為有效的頑固性，由於在從量子理論的繼承，在一個合理的答案的數學對象，必須執行的操作的數量組合增加。

字符串轉換在數學和物理，以及許多有價值的應用（例如，被稱為重寫長期的正式技術用在計算機代數系統領域）。

關於微笑的反應數據庫

在 2007 年，迅速 在過去一年的工作開始了一個人的審查的化學反應數據庫的組合，不久後開發的軟件支持圖像知識提取和蜘蛛終於實現了。 的微笑反應數據庫現在是 186.8 MB 的大小，它包含兩百萬對數千名受人尊敬的期刊和專利，從六個文件中提取反應的產品。 反應在每個數據庫文件的數據項出現連續行的文件，這是由換行符劃定。

獲得的微笑反應數據庫

[法律公告](#)

[價格信息表](#)



購買立即下載：

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#)

(選擇購買選項)

你可以下載最多的 3 倍，因此，請保存文件到一個可移動的光盤和存儲安全。

問題或評論？ [發電子郵件給我們](#)